SAND2000-2605C

"First Principles Applied to Software Safety - The Novel Use of Silicon Machinery"

Larry J. Dalton
Sandia National Laboratories
ljdalto@sandia.gov

Background

High Assurance Systems Engineering represents the attributes of 1) skilled/professional exercise of an engineering discipline(s) and 2) personal ethical commitment to the achievement of assurance objectives. Assurance objectives include all aspects of a product's life cycle and its operational profile, inclusive of physical environments it must operate in. Professional capability is a matter of academic preparation and acquired skills/experience. In Public Works projects, professional capability is measured through a Professional Engineer Certification process and is generally a legal contract requirement. Personal ethical commitment is not so easily measured and is analogous of the physicians Hippocratic oath. The dimension or strength of these attributes for any given product is set by the value propositions of the controlling business enterprise. Value propositions can be considered to be coefficients applied to the exercise of an engineering discipline and personal ethical commitment. The coefficients represent the 'thinking' of the business enterprise or responsible agency. In the case of the Department of Energy, value propositions for nuclear weapons represent the thinking that "*an accidental or unintended detonation of a nuclear weapon shall never occur.*" In the opposite direction, consumer products reflect business enterprise thinking such as 1) *time to market*, 2) *minimum product cost,* 3) *maximum profit*, and 4) *keep the shareholders happy*. In nuclear weapons, while resources are finite, safety is of utmost importance. A nuclear engineer's ethical creed that has been put forth is "a complete and pervasive intolerance to the compromise of safety." [Nick97] In contrast, by choice, the reliability, safety or security of a consumer product or service is often 'adjusted' (compromised) in order achieve the business enterprise value propositions. In many cases, the residual risk of product related losses and or the cost of litigation is mitigated through insurance vs. the cost of a better design. In most cases, the devotion to assurance of reliability, safety, and security is no stronger than the cost and schedule will allow as dictated by the associated value propositions. Assuring the reliability, safety and security of products then becomes a dichotomy.

The dichotomy is exacerbated by the growth in complexity of software and hardware and the propensity to use software in increasingly safety critical applications. We face tremendous challenges with current computational technologies, e.g. defensibly achieving 10^{-9} per hour likelihood of a catastrophic failure for the operational lifetime of all aircraft of one type. [FAA88, paragraphs 6.h(3) and 9.e(3)] Software is pervasive in engineered systems and it seems

DISCLAIMER

clear that after decades of research we have made little impact on how complex software systems are constructed in the trenches. It also seems clear that conventional testing methods, however carefully crafted, will not suffice in the future. A future where the size of software systems is far beyond human ability to analyze or assess from a safety perspective. In concert with the size of software is the continual growth in both capability and complexity of underlying digital hardware-computing platforms.

Overview

This paper presents a proposed methodology under investigation that may provide a radical new way of assuring the safety of software-based systems through a novel application of 'first principles' enabled by Micro-Electro-Mechanical-Systems (MEMS) technology, i.e. silicon machinery. 'First principles' as used herein is defined as theory that is defensible through fundamental laws of nature in the chemical, physical or mechanical structure of materials or assemblages thereof. The proposed methodology is limited to 'passive safety' as opposed to 'active safety' applications. *Passive safety* is defined herein to be a quality such that a potential hazard is mitigated (assured safe) by means that do not require action or energy to maintain, e.g. a mechanical 'stop block' designed to limit mechanical travel of a structure given design basis accident scenarios. *Active safety* is defined to be a system that requires action 1) to maintain a safe state or 2) to take the system to a safe state. Control of an operating nuclear reactor and an aircraft in flight are active safety system examples, i.e. active control is required to maintain or to reach a safe state.

The methodology proposed herein is based upon long standing safety principles employed in nuclear weapons. It is proposed that two of the long established nuclear weapon safety principles be applied to high consequence software systems. The nuclear weapon stronglink and the Unique Signal (UQS) concept are fundamental to nuclear weapon safety and represent the conceptual genesis of the approach taken herein [Spray91]. *Incompatibility* and *isolation* are the two fundamental nuclear weapon safety principles made possible by the nuclear weapon stronglink and UQS concepts.

In overly simplified terms, *Incompatibility* is made manifest by a discrimination function that unambiguously confirms intent to initiate a high consequence action. Intent can be of human or system origination. The UQS represents initiation intent and is communicated to a discriminator through a serial pulse train. Discrimination is performed mechanically; thereby eliminating many possible failure modes as would be present in software or digital hardware implementations (design/technology diversity). Only the UQS is *compatible* while the universe of all other signals is *incompatible*. The UQS is constructed in such a manner that makes it highly unlikely (e.g. to achieve desired safety of 10^{-9} per unit time) to be duplicated or simulated in normal environments and in a broad range of ill-defined abnormal environments.

Isolation in a nuclear weapon context is made manifest through the blockage of energy or data across/through a protected boundary until intent is unambiguously confirmed. In addition, in a nuclear weapon, *isolation* includes a safe response (preservation of *isolation*) to a collection of abnormal physical environments (physical threats).

The mapping of these well established safety concepts from nuclear weapons to complex software based systems is as follows: 1) the complex, high consequence software system must in-situ, in real-time, generate a 'key' (UQS corollary) which represents an unambiguous assertion of correct system behavior, 2) the 'key' must be discriminated confirming its *compatibility* or *incompatibility*, and 3) if the 'key' is deemed compatible, energy or data is allowed to pass across/through a protected boundary. In complex, high consequence software systems, this boundary can be considered either a physical or logical boundary, e.g. logical access to a block of memory or a physical barrier in an I/O line (mechanically controlled shutter in an optical signal path). The overarching objective is to be able to make defensible quantitative assertions about the safety of a system. The defensible quantitative assertions are based on the validity of abstracted models of behavior, the encoding of those models via mathematically based faithful execution vector generation, followed by mechanically based vector discrimination as a basis of optical path control. 'Valid' models of behavior must represent a sufficient number of observable events, states, parameters, sequencing, and/or timing relationships as specified by domain experts, which if determined to have been faithfully completed, would provide convincing evidence of system behavior as an enabling condition to the initiation of a high consequence event. MEMS technology is critical to the realization of isolation and incompatibility concepts in the context of a microscopic physical form, preserving essential mechanical attributes of the weapon stronglink, e.g. mechanical discrimination of the 'key' as well as the control of flow of energy or data. The notion of 'first principles' is captured in the pure mechanical construct of the MEMS device and its functions. That is, the behavior of the MEMS device can be evaluated strictly from a materials and mechanical construction/function viewpoint, devoid of execution of logic functions. The MEMS device provides the required mechanical discrimination as well as mechanically controlled optical air gap switches for control of flow of energy or data. The *isolation* attributes of the proposed MEMS device herein are limited to the control of information or energy flow across or through a protected boundary through a mechanical shutter function in an optical path. It does not (cannot) address a physical threat as in the nuclear weapon.

The concept presented here of encoding (faithful execution vector generation) models of behavior in-situ, in real-time, in conjunction with instantiation of isolation and incompatibility in the context of 'first principles' shifts the burden

away from exhaustive analysis and testing of 'what ifs' (the universe of possibilities) as a means of assurance. Instead, high consequence functions (potential hazards) are passively held in a safe state (isolated) until the precise specification of correct behavior is dynamically confirmed (compatibility) in-situ, in real-time. If behavioral models are constructed properly, one need not concern oneself with rare events or negative testing. Rare event and negative testing potentially become, in the proposed concept, only issues related to reliability, e.g. will the system function when we need it to. Properly constructed behavioral models should be sufficiently precise as to preclude the universe of possible failure conditions without having to explore that space. This places some analytical burden on ensuring that, for example, deviation from a state-chart instantiation of a behavioral model will be captured and manifest in an incorrect faithful execution vector and hence incompatibly. Research and development will be required for the theory, methods and tools to ensure that behavioral models possess the required attributes. A working version of a MEMS discrimination and controlled shutter device has been constructed and functionally evaluated. The existing working version has the capability to discriminate to 1 in 10^6 possible vectors. Further work in MEMS technology regarding integration of photonics is also required.

References

[FAA88] *System Design and Analysis*. Federal Aviation Administration, June 21, 1988. Advisory Circular 25.1309-1A

[Nick97] Quote from Bill Nickell, Director of Surety Assessment Center, Sandia National Laboratories at 1997 High Consequence Operations Safety Symposium

[Spray91] SAND91-1269 *The Unique Signal Concept for Detonation Safety in Nuclear Weapons*, Sandia National Laboratories, Stan Spray and J. Arlin Cooper

[Winter98] *Passive Safety in High-Consequence Systems*, Winter, Covan, Dalton, IEEE Computer, April 1998, Volume 31, Number 4, Pages 35-37

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.